基于药物和疾病特征关联的药物重定位混合推荐算法 *

刘 杰 a, b，金柳颀 b，景 波 a, b

(合肥工业大学 a. 工业与装备技术研究院; b. 国家“111 计划”老人福祉信息科技创新引智基地，合肥 230000)

摘 要：针对基于协同过滤的药物重定位算法进行了研究,考虑到数据稀疏性对协同过滤算法的巨大影响，提出一种基于药物和疾病特征关联的药物重定位混合推荐算法。该算法不仅使用了药物和疾病关系数据，还利用了药物结构、靶蛋白、副作用以及药物—疾病特征矩阵等信息计算药物之间的相似性，降低了数据稀疏性对推荐效果的影响，提高了推荐精度。经过对比实验发现，该算法具备较好的推荐效果，并能够发掘具有潜在联系的药物-疾病组合，从而进一步验证了该算法可以有效地应用于药物重定位。

关键词：药物重定位；数据稀疏性；疾病特征；混合推荐；相似度

中图分类号：TP301.6 doi: 10.3969/j.issn.1001-3695.2018.09.0634

Hybrid recommendation algorithm for drug repositioning based on association of drug and disease characteristics

Liu Jie[a,b], Jin Liuqi[b], Jing Bo[a,b]

(*a. Institute of Industry &Equipment Technology, b. National* "111 *Plan*" *Gerontechnology Innovate Base, HeFei University of Technology, Hefei* 230000, *China*)

Abstract: The algorithm of drug repositioning based on collaborative filtering was studied. Considering the great influence of data sparsity on collaborative filtering algorithm, this paper proposed a hybrid recommendation algorithm based on the association of drug and disease characteristics. The algorithm not only used the data of drug and disease, but also used the information of drug structure, target protein, side effect and drug-disease feature matrix to calculate the similarity between drugs, which reduced the influence of data sparsity to the recommendation effect and improves the precision of recommendation. The results of contrastive experiment showed that the algorithm has a good recommendation effect, and can explore the drug-disease combinations which have potential relationship, and further verified that the algorithm can be effectively applied to drug repositioning.

Key words: drug repositioning; data sparsity; disease characteristics; hybrid recommendation; similarity

0 引言

药物重定位，又称老药新用，指对于已批准应用于临床或者未上市但结构明确、生物活性已知的药品，通过进一步研究，扩大其适应症、发现其新的作用靶点[1]。

传统的新药研发通常要经历研究和开发两个阶段，每个阶段又有多个过程，是一个长期、艰难和昂贵的过程，尽管近年来药物研发的投入越来越高，但是新药的批准率却没有增加反而有降低的趋势。因此，能大大缩短药物研发所需的时间、经费且研发的成功率远远高于传统新药研发模式的“药物重定位”模式逐渐成为很多科研机构、医药企业看重的策略之一。

在考虑药物重定位时，可从不同角度开展研究，主要包括基于疾病、基于靶点和基于临床观察三个方面。但无论是从哪个方面进行药物重定位研究都需要研究人员对疾病和药物有全面和深刻的理解，研究人员将不同渠道来源的数据汇总成为大数据并加以总结和分析可以提高药物重定位的研究效率和成功率，但如何在海量数据中更为高效地发掘出有价值的信息也成为研究人员面对的一个重要问题，这就为个性化推荐算法应用于药物重定位提供了一个契机。

1 药物重定位算法

从药物重定位被提出以来，很多专家、学者投入大量精力研究药物重定位算法。Hopkins[2]提出网络药理学是药物发现的一个强有力的工具；Kinnings 等人[3]采用支持向量机（Support Vector Machines）改进了药物-靶标对接评估技术，并应用于寻找结核杆菌的直接抑制剂；Andronis 等人[4]提出利用文献挖掘方法结合大量生物学注释和可视化工具整合数据，有助于发现已有药物和新适应症之间的关系；Huang 等人[5]则将机器学习算法和拓扑图理论应用到非小细胞肺癌药物再定位中；文献[6,7]将推荐系统中的传统协同过滤算法改进后用于药物重定位中；Hu 等人[8]对约 7 000 个药物作用和疾病相关的基因表达谱进行了分析，依据 Pearson 相关系数计算表达谱之间的相似性，构建了包含 165 374 对药物—药物相似关系的网络，并采用聚类方法对药物进行分析。

由于数据稀疏性对传统协同过滤算法的推荐效果影响较大且用于计算药物之间相似度的药物—结构矩阵、药物—靶蛋白矩阵、药物—副作用矩阵以及药物—疾病矩阵的数据稀疏度都较高，所以将传统协同过滤算法应用于药物重定位虽然有一定的效果，但仍有不小的提升空间。本文提出一种基

收稿日期：2018-09-03；**修回日期：**2018-10-15 **基金项目：**安徽省 2017 年度重点研究与开发计划项目（1704e1002221）；国家高等学校学科创新引智计划资助项目（“111”）（B14025）

作者简介：刘杰（1992-），男，安徽青阳人，硕士研究生，主要研究方向为推荐算法研究、Web 前后端开发（hfutlj@outlook.com）；金柳颀（1991-），男，博士研究生，主要研究方向为人工智能、软件体系结构；景波（1994-），男，硕士研究生，主要研究方向为知识图谱、分布式实时系统.

于药物和疾病特征关联的药物重定位混合推荐算法，通过引入疾病特征向量和药物-疾病特征矩阵，在充分利用了相关数据矩阵的同时降低了数据稀疏性对推荐效果的影响，提高了基于个性化推荐的药物重定位算法的精确度。本文所提算法框架如图 1 所示。

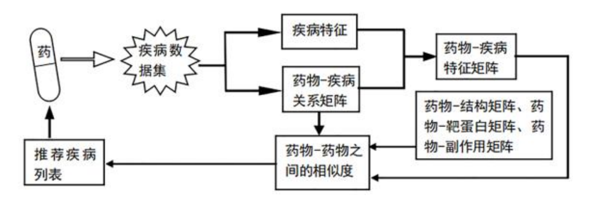


图 1　所提混合推荐算法架构

Fig. 1　Framework of the proposed hybrid recommendation algorithm

2　药物—疾病特征关系向量

传统协同过滤算法主要使用用户评分矩阵来计算用户之间的相似度，所以它的推荐效果很容易受到评分矩阵数据稀疏性的影响,很多学者研究此领域时都会考虑到解决数据稀疏性这一问题，其中就包括聚类和填值，文献[9]提出的一种正则化的局部学习方法就可以达到微阵列缺失值填补的效果。本文将个性化推荐算法应用于药物重定位领，将药物视为用户，疾病视为项目，药物—疾病关系矩阵看做评分矩阵，由于一种药物治疗的疾病类型有限，所以药物—疾病关系矩阵一般稀疏性都比较大,如果仅仅使用传统的协同过滤算法，往往很难有较好的推荐效果，考虑将疾病本身具有的属性引入到药物相似度计算当中，从而有更多的信息来计算药物相似度，减小数据稀疏性对推荐算法的影响。通过分析对疾病的描述，可以较为简单地获取疾病的一些属性，比如致病因子、患病性别、患病部位、发病过程等，这些属性会被用来描述疾病的特征从而用于药物相似度计算。

用 $D=\{d_1,d_2,...,d_n\}$ 来表示疾病集合，那么每种疾病的特征值可以表示为 $F_{d_j}=a_{d_j1}a_{d_j2}...a_{d_jl}$，其中 a_{d_jk} 表示疾病 d_j 属性 k 的类别，F_{d_j} 表示疾病 d_j 的特征，l 表示用来表示疾病特征的属性个数。

有了每种疾病的特征之后，可以开始计算药物—疾病特征关系向量，用 $Y=\{y_1,y_2,...,y_m\}$ 表示药物集合，用 $C=\{c_1,c_2,...,c_q\}$ 表示所有疾病属性类别集合，其中 q 表示疾病属性类别总数，则可以用 $F_{y_j}=\{w_{y_j1},w_{y_j2},...,w_{y_jq}\}$ 表示药物 y_j 的药物-疾病特征关系向量，其中 w_{y_jk} 反映药物 y_j 对疾病属性类别 c_k 的作用程度。对药物 y_j 的药物-疾病特征关系向量可以通过以下步骤计算：

获取药物 y_j 有疗效的疾病集合 $D_{y_j}=\{d_{j1},d_{j2},...,d_{jz}\}$，z 表示药物 y_j 有疗效的疾病个数。计算集合 D_{y_j} 中每种疾病特征值 $F_{d_{ji}}$，然后计算出 D_{y_j} 总特征值 $F_{D_{y_j}}$ 为

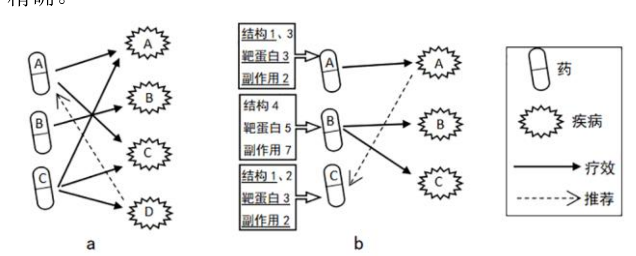
$$F_{D_{y_j}}=\sum_{i=1}^{h}F_{d_{jk}} \tag{1}$$

统计 $F_{D_{y_j}}$ 中每种属性类别的个数，按类别填入药物 y_j 的药物—疾病特征关系向量 $F_{y_j}=\{w_{y_j1},w_{y_j2},...,w_{y_jq}\}$。

3　基于药物—疾病特征矩阵的混合推荐

3.1　基本原理

近年来，个性化推荐系统在影视、电子商务、网络广告、社交网络等领域都得到了巨大发展[10]，而作为推荐系统核心的推荐算法也成为越来越多专家学者的研究对象并被灵活应用于其他场景，如杨矫云等将推荐算法应用于选择生物块以加速遗传电路设计[11]。个性化推荐算法种类很多，比如协同过滤算法、基于内容的推荐以及基于人口统计学的推荐等，本文所提算法主要结合协同过滤算法和基于人口统计学的推荐算法，其中协同过滤算法从被提出开始就被应用于各个场景和领域，从一开始在邮件系统中的应用到后面新闻、电影、音乐以及应用最广泛的电子商务领域，在早期协同过滤算法的研究中，Sarwar 等人[12]将基于用户和基于项目的协同过滤算法进行了对比研究发现基于项目的协同过滤具有更好的性能；而最著名的电子商务推荐系统应属亚马逊网络书店，Linden 等人[13]详细描述了亚马逊网站中使用的协同过滤算法的原理，并与传统的协同过滤算法对比说明亚马逊推荐算法对的优势，进一步推动了协同过滤算法的发展。然而协同过滤算法在具有很多优势的同时也存在冷启动和数据稀疏性等问题，很多研究者都将精力投入到如何弥补缺点从而提高推荐效果的研究中。基于人口统计学的推荐就是根据人口统计学数据对每个用户聚类，通过聚类计算用户相似度[14]。本文所提出的算法主要应用于药物重定位，通过将药物看做用户，疾病看做项目，融合基于人口统计学推荐和基于用户的协同过滤算法的基本思想，并结合药物-疾病特征矩阵，从而产生本文所提混合推荐算法。将基于人口统计学推荐和基于用户的协同过滤算法应用于药物重定位的工作原理如图 2 所示。图中，(a)是基于用户的协同过滤算法原理示意图，药物 A 对疾病 A、C 有疗效，药物 C 对疾病 A、C、D 有疗效，此时认为药物 A、C 相似，因此认为药物 A 对疾病 D 也有疗效，因此将疾病 D 推荐给药物 A，此种算法中用来衡量药物之间相似度的就是药物-疾病矩阵；(b)是基于人口统计学的推荐算法原理示意图，由图可以看出，药物 A 和药物 C 具有相似的结构、相同的靶蛋白和相同的副作用，此时认为药物 A、C 相似，因此将药物 A 有疗效的疾病 A 推荐给药物 C，此时将药物-结构矩阵、药物-靶蛋白矩阵以及药物-副作用矩阵看成药物的人口统计学数据，并利用这三种数据对药物聚类计算药物相似度，因为矩阵数据比单纯的某种性质数据更加丰富，所以利用这三种矩阵来衡量药物之间的相似度更加精确。



(a)与基于人口统计学的推荐算法　　(b)原理图

图 2　基于用户的协同过滤算法

Fig. 2　User based collaborative filtering algorithm(a) and demographic-based recommendation(b)

在按步骤计算得到每种药物的药物-疾病特征关系向量后，可以将它们组成所需的药物-疾病特征矩阵，基于药物—疾病特征矩阵并融合上述基于人口统计学推荐和基于用户的协同过滤算法，联合计算出药物之间的相似度并作出推荐。

3.2　相似度计算

本文提出的混合算法包含三部分相似度计算：a)基于药物—结构、药物—靶蛋白和药物—副作用矩阵的相似度；b)基于药物—疾病矩阵的相似度；c)基于药物—疾病特征矩阵的相似度。用 Sim_D 表示相似度 a，Sim_C 表示相似度 b，Sim_F 表示相似度 c，总相似度表示为 Sim_T。因此，药物 y_i 与药物 y_j 之间的总相似度可以用式（2）表示。

$$Sim_T(y_i,y_j)=\alpha Sim_C(y_i,y_j)+\beta Sim_F(y_i,y_j)+(1-\alpha-\beta)Sim_D(y_i,y_j) \tag{2}$$

其中 $0\le\alpha,\beta,\alpha+\beta\le1$；相似度 $Sim_C(y_i,y_j)$ 可以用余弦相似度计

算，如式（3）所示。

$$Sim_C(y_i,y_j)=\frac{\sum_{l=1}^{n}r_{y_il}\cdot r_{y_jl}}{\sqrt{\sum_{l=1}^{n}r_{y_il}^2}\sqrt{\sum_{l=1}^{n}r_{y_jl}^2}} \tag{3}$$

其中：n 表示疾病总数，r_{y_hl} 代表药物 y_h 对疾病 d_l 是否有效。

$$r_{y_hl}=\begin{cases}1,\ y_h\text{对}d_l\text{有效}\\0,\ y_h\text{对}d_l\text{无效}\end{cases} \tag{4}$$

其中 h 可以取 i 或者 j。

另外，相似度 $Sim_C(y_i,y_j)$ 可以用式（5）计算。

$$Sim_C(y_i,y_j)=\frac{E_{y_i}\cap E_{y_j}}{E_{y_i}\cup E_{y_j}} \tag{5}$$

其中：E_{y_i} 表示药物 y_i 有疗效的疾病集合，E_{y_j} 表示药物 y_j 有疗效的疾病集合；考虑到药物-疾病矩阵的特殊性并经过实验验证后发现，采用式（5）计算药物之间的相似度推荐精度更高，所以本文所提算法采用式（5）计算相似度 $Sim_C(y_i,y_j)$。

相似度 $Sim_D(y_i,y_j)$ 主要包含三个部分，基于药物-结构矩阵相似度、基于药物-靶蛋白矩阵相似度和基于药物-副作用矩阵的相似度，但本文认为药物-结构矩阵和药物-靶蛋白矩阵贡献的相似度程度相同，所以 $Sim_D(y_i,y_j)$ 可以表示如下：

$$Sim_D(y_i,y_j)=\gamma\cdot Sim_E(y_i,y_j)+\frac{1-\gamma}{2}\cdot(Sim_S(y_i,y_j)+Sim_P(y_i,y_j)) \tag{6}$$

其中：$0\le\gamma\le1$，$Sim_E(y_i,y_j)$ 表示基于药物-副作用矩阵的药物 y_i，y_j 的相似度，$Sim_S(y_i,y_j)$ 表示基于药物-结构矩阵相似度、$Sim_P(y_i,y_j)$ 表示基于药物-靶蛋白矩阵相似度，与 $Sim_C(y_i,y_j)$ 类似，这三种相似度也用式（3）中方法或者式（5）中方法来计算，综合考虑比较后，本文所述算法选择式（5）所述方法来计算这三种相似度。

由于药物—疾病特征矩阵的特殊性，相似度 $Sim_F(y_i,y_j)$ 的计算并不采用余弦相似度计算，而是通过式（7）来计算。

$$Sim_F(y_i,y_j)=\sum_{l=1}^{q}\min(w_{y_il},w_{y_jl})\div\sum_{l=1}^{q}\max(w_{y_il},w_{y_jl}) \tag{7}$$

其中：q 表示疾病属性类别总数，w_{y_il} 反映药物 y_i 对疾病属性类别 c_l 的作用程度，w_{y_jl} 反映药物 y_j 对疾病属性类别 c_l 的作用程度。

3.3 评分表示方法

得出药物 y_i 与其他所有药物的相似度后，选取相似度值按大小排名前 t 个药物作为药物 y_i 的邻居用来计算药物 y_i 对各种疾病的有效值，药物 y_i 对疾病 d_l 的有效值 r_{y_il} 可以用式（8）计算。

$$r_{y_il}=\frac{\sum_{y_k\in N_{y_i}}Sim_T(y_i,y_k)\cdot r_{y_kl}}{\sum_{y_k\in N_{y_i}}Sim_T(y_i,y_k)} \tag{8}$$

其中：$N_{y_i}=\{y_1,\ldots,y_t\}$ 表示药物 y_i 的邻居，药物 y_k 是其中一个邻居。

4 实验结果分析

本文实验数据集以文献[6]中所使用到的相关数据信息为基础，并基于这些数据做了一些处理。为了充分利用相关数据并整理出本文所需数据，首先从药物—疾病、药物—结构、药物—靶蛋白以及药物—副作用表格数据提取出共有的药物，然后用这些共有药物的四种矩阵数据来开展本文算法验证实验。本文所提算法需要用到的药物-疾病特征矩阵数据需要基于疾病特征获取，本文所涉及疾病特征均为人工查询、整理和标注。数据处理之后用于本实验的药物总数为 536 种药物，疾病为 720 种、1386 种副作用、776 种靶蛋白和 882 种结构，其中训练集占 90%，训练集药物—疾病矩阵数据稀疏程度为 0.9944，药物—副作用矩阵数据稀疏程度为 0.9451，药物—靶蛋白矩阵数据稀疏程度为 0.9956 以及药物—结构矩阵数据稀疏程度为 0.8603（稀疏度的计算主要是根据数据矩阵的无作用数据占总数据的比例）。在疾病特征方面，本文采用人工查询、整理和标注的方式对涉及的 720 种疾病的特征进行了梳理，并主要划分为四个属性：病因（包括细菌感染、真菌感染、病毒感染以及寄生虫等），发病性别（男，女，儿童），病症部位（心脏，四肢，消化系统，神经系统以及皮肤等），发病过程（急性、慢性）。

4.1 评测指标

为表明实验结果好坏，本文首先采用了平均绝对误差（mean absolute error, MAE）作为度量算法优劣的指标。

$$MAE=\frac{1}{N}\sum_{y_i\in Y_p}\sum_{d_j\in D_i}\left|\overline{r_{ij}}-r_{ij}\right| \tag{9}$$

其中：Y_p 表示待预测药物集合，D_i 表示药物 y_i 有预测值得疾病集合，N 表示所有预测值的总数，$\overline{r_{ij}}$ 表示药物 y_i 对疾病 d_j 的预测值，r_{ij} 表示药物 y_i 对疾病 d_j 的真实值。

然而通过进一步研究发现，根据邻居药物对疾病的有效性以及与邻居之间的相似度计算出目标药物对各种疾病的预测值，如果如传统的电影评分推荐系统那样，用这个预测值本身表示药物对该疾病的有效性，然后计算 MAE，此时 MAE 能够反映算法优劣。但由于在药物-疾病矩阵这一背景下，药物对疾病有效就表示为 1，无效就表示为 0，正是由于存在这一特点，在本实验中仅用 MAE 并不能很精确地体现各种算法性能优劣，此时更合适的是使用召回率(recall)、准确率(precision)以及 F 值这三个指标来衡量算法优劣，召回率、准确率以及 F 值的定义如（10）~（12）所示。

$$Recall=\frac{N_P}{D_T} \tag{10}$$

$$Precision=\frac{N_P}{N_T} \tag{11}$$

$$F=\frac{2\cdot Recall\cdot Precision}{Recall+Precision} \tag{12}$$

其中：N_P 表示推荐列表中评分与真实情况相符的项目数，D_T 表示用户真实评分过的项目总数，N_T 表示推荐列表中的项目总数，即推荐列表长度。理论上，当 precision 越高，同时 recall 也越高，算法越好，但事实上这两者在某些情况下有矛盾的，而 f 值是 precision 和 recall 加权调和平均，也就是一个综合评价指标。

4.2 实验结果

为了验证仅使用 MAE 作为评测指标并不能很好地反映算法优劣，本文首先仅用 MAE 作为评测指标进行了一组实验，因为本文所提算法涉及多种相似度计算，总相似度由式（2）计算得出，由式（2）可知，如果要计算总相似度，则需要决定因子 α 和 β 的取值，经过多次实验后最终确定 α =0.6、β =0.3 以及 $\gamma=0.8$，同时为了更直观地体现本算法在性能上的提升，本文采用对比实验，实验结果如图 3 所示，其中横轴 k-near 表示推荐过程中邻居个数。

通过图 3 可以看出，如果采用 MAE 衡量算法性能优劣，基于药物本身特点的推荐和本文所提算法的实验结果都优于

传统协同过滤算法，而且随着邻居数的增加，效果越来越好，最后趋于收敛。但从图 3 也可以看出基于药物本身特点的推荐效果与本文所提算法相比效果十分接近，而它的复杂程度却远远小于本文所提算法，所以这并不是一个好的实验结果，由此引发对实验方案和评测指标的进一步思考。

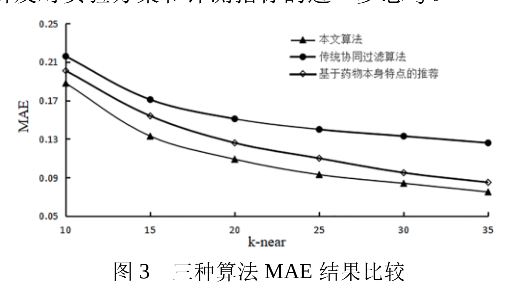


图 3　三种算法 MAE 结果比较

Fig. 3　MAE results' comparison of the three algorithms

为了进一步优化实验，本文使用召回率、准确率以及 F 值来作为评测指标进行实验，考虑到药物—疾病矩阵的特点：药物对疾病有效就表示为 1，无效就表示为 0，本文提出另一种确定预测值的方案，从而解决算法效果不佳以及 MAE 不能较为精确的反映算法效果的问题。通过相似度 Sim_r 计算得到的预测值，往往是介于 0 到 1 之间的小数，而药物与疾病之间只存在两种状态：有效或者无效，如果像类似于电影推荐系统那样，就以计算得到的预测值作为药物与疾病的状态值，显然是不合理的，因此本文确定阈值为 λ，将计算得出的预测值大于 λ 的置为 1，小于或等于 λ 的预测值置为 0，以此计算 MAE，同时，只向药物推荐预测值为 1 的疾病。这样处理之后显然会存在一个问题，那就是召回率会降低。此时将召回率和准确率带入到本文所在场景中：用一种方法对一种药物进行一系列分析，得出了 20 种这种药物可能有疗效的疾病，然后发现此药物真的对这 20 种疾病都有效，但实际上这个药物对 40 种疾病都有疗效。在上述场景中，准确率是 100%，召回率是 50%，但这种结果感觉上是很可靠的，因为它能够很准确的发现药物的疗效，虽然它并不能发掘出所有药物能够治疗的疾病，但是只要它发掘出一种疾病，这个疾病在很大概率上是能被相应药物治疗的，因此准确率越大越符合药物重定位的思想，但是如果在保证准确率很大的前提下仍然能有较好的召回率，即推荐的疾病不仅准而且全，那效果当然更好。

基于上述分析，本文最终采用按初步预测值是否超过阈值置最终预测值为 1 或 0 且仅推荐预测值为 1 的方法来进行推荐，并且在实验之后确定 λ 为 0.7 时具有最佳效果，实验结果如图 4~6 所示，可以看出，图中没有呈现基于药物特点推荐的实验结果，但实际上基于药物特点推荐的实验结果也有，只不过该种算法结果与本文所提算法和传统协同过滤算法差距悬殊，所以就并没有放在一起比较，而是以表格形式呈现了结果，基于药物特点推荐的实验结果如表 1 所示。

由图 4~6 可以看出，本文所提算法在各个评测指标上的实验结果均优于传统协同过滤算法，在具有较高准确率的同时也能保证较好的 F 值。同时也可以看出，仅仅基于药物—结构、药物—靶蛋白和药物-副作用矩阵进行推荐的效果并不理想，这也进一步证明了开始用未优化的 MAE 来衡量算法优劣的不合理性。而传统的协同过滤算法在性能上虽然不如本文所提算法，但仍然具有不错的效果，这也进一步表明个性化推荐算法在药物重定向领域应用具有可行性。

本文所提算法验证实验在计算准确率的同时还记录了推荐列表中与药物-疾病矩阵中实际情况不符的药物-疾病组合，通过分析这些组合可以进一步验证本文所提算法在药物开发方面能否起到一定的作用，部分相关药物-疾病组合具体情况如表 2 所示。

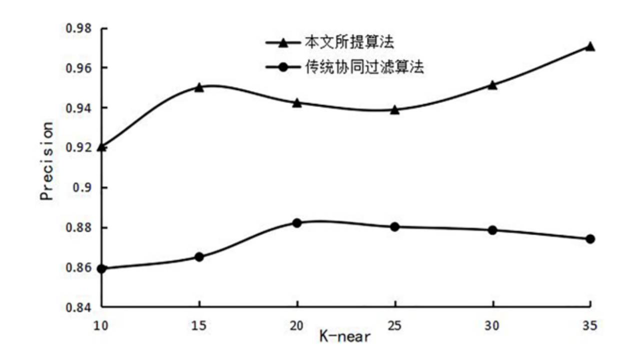


图 4　本文所提算法与传统协同过滤算法准确率对比

Fig. 4　Comparison of precision between the proposed algorithm and the traditional collaborative filtering algorithm

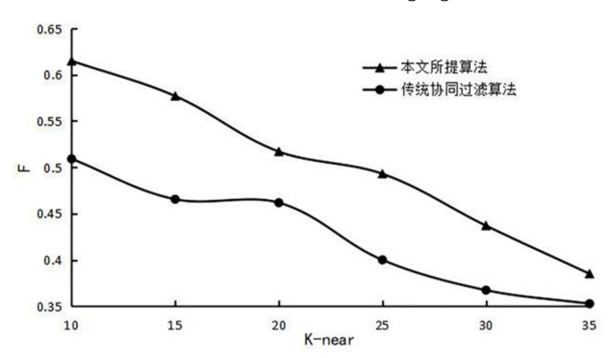


图 5　本文所提算法与传统协同过滤算法 F 值对比

Fig. 5　Comparison of F-measure between the proposed algorithm and the traditional collaborative filtering algorithm

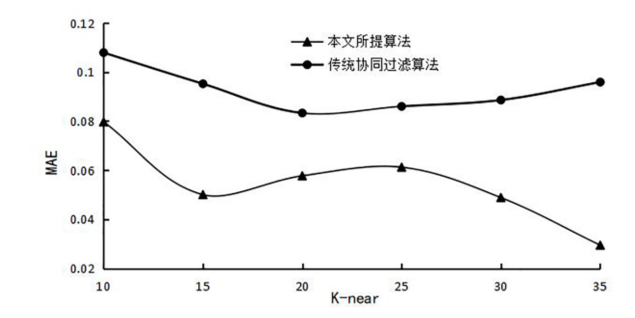


图 6　本文所提算法与传统协同过滤算法 MAE 对比

Fig. 6　Comparison of MAE between the proposed algorithm and the traditional collaborative filtering algorithm

表 1　基于药物特点推荐的实验结果

Table 1　Experimental results recommended based on drug characteristics

K-near	10	15	20	25
MAE	0.1538	0.1212	0.2857	0
Recall	0.2	0.1054	0.0182	0.0036
Precision	0.8462	0.8788	0.7143	1
F	0.324	0.188		

从表 2 中可以看出，预测有疗效的药物—疾病组合与实际情况相符，其中除了比较明显有疗效的药物—疾病组合外，头孢曲—嗜血杆菌感染组合的疗效可以在文献[15]中的实验得到证实；葡萄球菌感染多表现为皮肤、软组织感染，也可导致病情严重、危及生命的败血症、肺炎、脑膜炎等，此外尚可引起异物相关感染、尿路感染，这与头孢泊肟有疗效的

几种感染相符；卓—艾综合征的临床表现包括：消化性溃疡、呕吐、腹泻等，其中呕吐症状与胃复安疗效相符。另外，革兰氏染色阴性杆菌对克林霉素耐药性好，而本文所提算法却预测了克林霉素对革兰氏染色阴性杆菌有疗效，这可能是由于某些疾病特征选取不合理造成的，这也是后期需要进一步研究、优化的地方。所以综上可以看出，本文所提算法确实可以发现一些现实中有治疗关系的药物—疾病组合，而且精度较高，进一步表明本文所提推荐算法应用于药物重定位的可行性和有效性，因此本文所提算法得出的推荐结果理论上可以在药物开发中起到辅助作用。

表 2 预测药物-疾病组合以及药物实际用途

Table 2 Drug-disease combinations predicted and the practical uses of the drugs

药物名称	预测有疗效的疾病	药物实际用途
氟卡胺	室上颤动	用于多种心律失常的防治,如室上性心动过速、心房颤动、单源性及多源性室性过早搏动综合征及其他抗心律失常药无效的病例
卡托普利	高血压	被应用于治疗高血压和某些类型的充血性心力衰竭
头孢曲松	嗜血杆菌感染	治疗呼吸道感染、泌尿系统感染、淋病
头孢泊肟	葡萄球菌感染	适用于敏感菌所致的支气管炎、肺炎及泌尿系统、皮肤和软组织、中耳、扁桃体等部位的感染
胃复安	卓-艾综合征	止吐药，可用于术后以及药物所引起的呕吐以及对胃胀气性消化不良、恶心、呕吐也有较好的疗效
喹硫平	精神分裂症	也可以减轻与精神分裂症有关的情感症状如抑郁、焦虑及认知缺陷症状
醋丁洛尔	室上性心动过速	由于阻滞心脏起搏点电位的肾上腺素能兴奋故用于治疗心律失常
碳酸锂片	惊恐障碍	主要治疗躁狂症，对躁狂和抑郁交替发作的双相情感性精神障碍有很好的治疗和预防复发作用
克林霉素	大肠杆菌感染/嗜血菌感染	临床上主要用于厌氧菌引起的腹腔和妇科感染，是金黄色葡萄球菌骨髓炎首选治疗药物

5 结束语

药物重定位是药物研发的一种重要策略，良好的药物重定位策略对于医疗健康领域有着重要意义。本文将个性化推荐算法应用到药物重定位，通过关联药物与疾病特征，同时联系药物—疾病关联矩阵、药物—结构、药物—靶蛋白和药物—副作用关联矩阵，设计了一种基于药物和疾病特征关联的药物重定位混合推荐算法，实验结果表明本文所提算法在性能上优于传统协同过滤算法等其他推荐算法，并能够有效地应用于药物重定位中。

参考文献：

[1] Chong C R, Sullivan D. New uses for old drugs. Nature[J]. Nature, 2007, 448(7154): 645-646.

[2] Hopkins A L. Network pharmacology: the next paradigm in drug discovery[J]. Nature Chemical Biology, 2008, 4(11): 682-690.

[3] Kinnings S L, Liu Nina, Tonge P J, *et al.* A machine learning-based method to improve docking scoring functions and its application to drug repurposing[J]. Journal of Chemical Information & Modeling, 2011, 51(5): 1195-7.

[4] Andronis C, Sharma A, Virvilis V, *et al.* Literature mining, ontologies and information visualization for drug repurposing[J]. Briefings in Bioinformatics, 2011, 12(4): 357.

[5] Huang C H, Chang M H P, Hsu C W, *et al.* Drug repositioning for non-small cell lung cancer by using machine learning algorithms and topological graph theory[J]. BMC Bioinformatics, 2016, 17(Suppl 1): 2.

[6] 林耀进，张佳，林梦雷，等.基于协同过滤的药物重定位算法[J].南京大学学报:自然科学版，2015，51(4): 834-841. (Li Yaojin, Zhang Jia, Lin Menglei, *et al.* Drug repositioning algorithm based on collaborative filtering [J]. Journal of Nanjing University：Natural Science, 2015, 51 (4): 834-841.)

[7] 章啸. 协同过滤算法在药物重定位中的研究与应用[D]. 上海：东华大学,2017. (Zhang xiao. Research and application of collaborative filtering algorithm for drug repositioning [D].Shanghai:Donghua University, 2017.)

[8] Hu Guanghui, Agarwal P. Human disease-drug network based on genomic expression profiles[J]. PLOS One, 2009, 4(8): e6536.

[9] Wang Aiguo, Chen Ye, An Ning, *et al.* Microarray missing value imputation: a regularized local learning method[J]. IEEE/ACM Trans on Computational Biology & Bioinformatics, 2018, PP (99).

[10] 冯蓓蓓.个性化推荐系统综述 [J].科技展望，2017，27(12). (Feng Beibei. Survey of personalized recommendation system [J]. Science and Technology, 2017, 27(12) .)

[11] Yang Jiaoyun, Song Yu, Gong Bowen, *et al.* Biobrick chain recommendations for genetic circuit design [J]. Computers in Biology & Medicine, 2017, 86: 31-39.

[12] Sarwar B, Karypis G, Konstan J, *et al.* Item-based collaborative filtering recommendation algorithms [C]//Proc of International Conference on World Wide Web.New York:ACM Press,2001:285-295.

[13] Linden G, Smith B, York J. Amazon. com recommendations: item-to-item collaborative filtering [J]. IEEE Internet Computing, 2003, 7(1): 76-80.

[14] Li Guo, Zhang Zhibin, Liu Fangxian, *et al.* Nonlinear combinatorial collaborative filtering recommendation algorithm[J]. Journal of Computer Applications, 2011, 31(11): 3063-3067.

[15] 陆权，张泓，季芳,等.头孢曲松对肺炎链球菌和流感嗜血杆菌体外抗菌活性的研究 [J]. 临床儿科杂志，2001，19(5): 306-307. (Lu Quan, Zhang Hong, Ji Fang, *et al.* Antibacterial activity of ceftriaxone to streptococcus pneumonia and hemophilus influenza in vitro test [J]. Journal of Clinical Pediatrics,2001, 19(5):306-307.)